

Probabilistic Interpretation for Correntropy with Complex Data

João P. F. Guimarães Aluisio I. R. Fontes Joilson B. A. Rego Allan de M. Martins
 joao.guimaraes@ifrn.edu.br aluisio.rego@ifrn.edu.br jotarego@gmail.com allan@dee.ufrr.br

Abstract

Recent studies have demonstrated that correntropy is an efficient tool for analyzing higher-order statistical moments in nonGaussian noise environments. Although it has been used with complex data, some adaptations were then necessary without deriving a generic form so that similarities between complex random variables can be aggregated. This paper presents a novel probabilistic interpretation for correntropy using complex-valued data called complex correntropy. An analytical recursive solution for the maximum complex correntropy criterion (MCCC) is introduced as based on the fixed-point solution. This technique is applied to a simple system identification case study, as the results demonstrate prominent advantages regarding the proposed cost function if compared to the complex recursive least squares (RLS) algorithm. By using such probabilistic interpretation, correntropy can be applied to solve several problems involving complex data in a more straightforward way.

Index Terms

complex-valued data correntropy, maximum complex correntropy criterion, fixed-point algorithm.

1. Introduction

Defining the relationship between the input and output signals in a given system is a common problem widely found in distinct engineering areas [1], [2], [3], [4]. The classic regression solution is then extensively adopted using the mean square error (MSE) as a cost function in order to minimize the error between the input signal and the desired output. However, many authors have developed methods based on correntropy as a cost function in the last few years since such approach improves the fitting performance in nonGaussian noise environments [5], [6], [7].

Correntropy is a similarity measure between two variables, which contains information from all the even statistical moments, being a generalization of the correlation concept [8]. Several techniques have proposed the use of correntropy in adaptive system training, thus demonstrating its excellent performance in practical applications such as noise cancellation in speech signals [9], system identification [10], [11], [12], [13], and face recognition [14], where the errors are typically nonGaussian.

On the other hand, many cases typically involve a processed signal that belongs to the complex domain, as the complex RLS algorithm is widely used as a possible solution [15]. However, few studies have explored the use of correntropy as a cost function in problems involving complex-valued data. The work developed in [16] presents a complex-valued blind equalization algorithm for quadrature amplitude modulation (QAM) and complex channel environments based on the correntropy criterion. The study is motivated by the improved performance achieved by information theoretic learning (ITL) methods when compared to MSE-based approaches. A robust adaptive carrier frequency offset (CFO) algorithm was introduced in [17] for orthogonal frequency division multiplexing (OFDM) purposes, which also deals with QAM and phase-shift keying (PSK) complex symbols. However, correntropy as applied to complex-valued data has not yet been properly formalized.

Within this context, this paper presents a new probabilistic interpretation for correntropy with complex-valued data, which is defined as complex correntropy. It is based on the probability function in multidimensional spaces using the Parzen estimator. In addition, a novel analytical recursive solution to the MCCC applied as a cost function based on the fixed-point solution is proposed. The results demonstrate the advantages of the proposed cost function in nonGaussian environments when compared to the RLS algorithm for noise cancellation purposes. The remaining sections of this paper are organized as follows. Section II reviews

the probabilistic interpretation of correntropy and extends this concept to complex-valued data. Section III presents a closed form recursive solution to MCCC. Simulation results are presented in section IV, while a proper comparative analysis with the RLS algorithm performance is presented. Finally, relevant conclusions are given in Section V.

2. Probabilistic Interpretation of Correntropy

This section reviews the probabilistic interpretation of correntropy applied to real-valued data, so that it can be further extended to complex-valued data.

2.1. Correntropy Applied to Real-Valued Data

Correntropy is directly related to the probability regarding how similar two random variables are. In fact, correntropy is the exact estimate of such probability when a Parzen estimator is used for the joint probability [18]. Firstly, let us consider two arbitrary scalar random variables X and Y , as the correntropy between them is defined as:

$$V(X, Y) = \hat{P}(X = Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XY}(x, y) \delta(x - y) dx dy \quad (1)$$

In most cases, the real distribution is unknown and only a finite number of data samples $(x_n, y_n), n = 1, 2, \dots, N$ is available. However, it is possible to use the L-dimensional Parzen estimation with a Gaussian kernel to obtain $f_{XY}(x, y)$ as:

$$\hat{f}_{X_1, X_2, \dots, X_L}(x^1, x^2, \dots, x^L) = \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^L G_{\sigma}(x^l - x_n^l) \quad (2)$$

where $G_{\sigma}(x)$ is defined as

$$G_{\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3)$$

Notation x_n^l represents the n -th data sample for the l -th component of the L-dimensional random vector while σ is the kernel bandwidth, also known as the kernel size. In order to define correntropy for the real domain, the work presented in [19] considers L=2 in equation (2):

$$\hat{f}_{XY}(x, y) = \frac{1}{N} \sum_{n=1}^N G_{\sigma}((x - x_n) + (y - y_n)) \quad (4)$$

Substituting (4) in (1) gives:

$$V(x, y) = \int \int \frac{1}{N} \sum_{n=1}^N G_{\sigma}((x - x_n) + (y - y_n)) \delta(x - y) dx dy \quad (5)$$

If $x = y$, equation (5) can be rewritten as:

$$V(X, Y) = \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}((x - x_n) + (y - y_n)) du \Big|_{x=y=u}$$

$$V(X, Y) = \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}((u - x_n)) G_{\sigma}(u - y_n) du \quad (6)$$

where u represents the value assumed by x and y over the line $x = y$.

Equation (6) can be solved as

$$V(X, Y) = \frac{1}{N} \sum_{n=1}^N G_{\sigma}(x_n - y_n) \quad (7)$$

which corresponds to the expression that represents correntropy when applied to real-valued random variables X and Y [18].

2.2. Correntropy Applied to Complex-Valued Data

Statistical signal processing in the complex domain has traditionally been viewed as a straightforward extension of the corresponding algorithms in the real domain [16]. This paper is then supposed to present a probabilistic interpretation based on Parzen estimator defined according to equation (2) to measure the similarity between two complex variables. Assuming two random complex variables $C_1 = X + jZ$ and $C_2 = Y + jS$, where $C_1, C_2 \in \mathbb{C}$, and X, Y, Z, S are real-valued random variables, it is possible to use correntropy to measure the probability for which such complex numbers are equal. For this purpose, the correntropy concept must be extended for more than two variables, what can be performed when assuming that the probability regarding $C_1 = C_2$ causes the respective real and imaginary parts of C_1 and C_2 to be the same. Such probability can be stated in the form:

$$P(C_1 = C_2) = P(X = Y \text{ and } Z = S) \quad (8)$$

By using correntropy, such probability can be estimated as:

$$V(C_1, C_2) = \hat{P}(C_1 = C_2) \quad (9)$$

The probability interpretation of correntropy can be used to estimate the joint probability density f_{XYZS} as:

$$\begin{aligned}
V(C_1, C_2) &= \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) \delta(x-y) \delta(z-s) dx dy dz ds
\end{aligned} \quad (10)$$

If $x = y$ and $z = s$, equation (10) can be rewritten as:

$$V(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2}$$

$$V(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 \quad (11)$$

It is then possible to replace \hat{f}_{XYZS} for the Parzen estimator defined in equation (2) using $L = 4$:

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}(x - x_n) \cdot G_{\sigma}(y - y_n) + \\
&G_{\sigma}(z - z_n) \cdot G_{\sigma}(s - s_n) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_{\sigma}(u_1 - x_n) G_{\sigma}(u_1 - y_n) + \\
&G_{\sigma}(u_2 - z_n) G_{\sigma}(u_2 - s_n) du_1 du_2 \\
&= \frac{1}{N} \sum_{n=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_{\sigma}(u_1 - x_n) G_{\sigma}(u_1 - y_n) + \\
&G_{\sigma}(u_2 - z_n) G_{\sigma}(u_2 - s_n) du_1 du_2
\end{aligned} \quad (12)$$

Solving the double integral in (12) gives:

$$V(C_1, C_2) = \frac{1}{N} \sum_{n=1}^N G_{\sigma\sqrt{2}}(x_n - y_n) G_{\sigma\sqrt{2}}(z_n - s_n) \quad (13)$$

Equation (13) is then defined as correntropy for two complex random variables or simply complex correntropy. There are no assumptions or restrictions for its application to generic data e.g. constant modulus or argument, since it represents a complete measure of similarity between two random variables. It is important to understand the effect of the estimator when computing correntropy as a probability estimation, considering that the case where the imaginary part is equal to zero in both random variables could be misinterpreted. As a result of the Parzen estimator effect, it does not lead to correntropy as defined in equation (7), although the same result can be obtained by only adjusting the kernel size. This is the reason why real-valued correntropy does not add up to 1 if $x = y$. Equation (13) can also be further analyzed according to its respective Taylor series expansion. In

addition, it is possible to write the average sum as the expected value in the Parzen estimator, which leads to:

$$V(C_1, C_2) = \frac{1}{2\pi\sigma^2} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m \sigma^{2m} n!} E[(x-y)^{2m} + (z-s)^{2m}]$$

$$V(C_1, C_2) = \frac{1}{2\pi\sigma^2} + \frac{k_1}{\sigma^4} E[(C_1, C_2)(C_1, C_2)^* + h_{\sigma}(C_1, C_2)] \quad (14)$$

considering that

$$h_{\sigma}(C_1, C_2) = \frac{1}{2\pi\sigma^2} \sum_{m=2}^{\infty} \frac{(-1)^m}{2^m \sigma^{2m} n!} (E[(x-y)]^{2m} + E[(z-s)]^{2m}) \quad (15)$$

where $h_{\sigma}(C_1, C_2)$ is a term that contains all higher-order moments, whose components in the denominator depend on σ considering that the first term includes σ^6 .

According to equation (14), the higher-order terms represented by h_{σ} tend to zero faster than the second term as σ increases. It is worth to mention that the second term corresponds exactly to the covariance involving two complex variables C_1 and C_2 . Hence, as the kernel size increases, the complex correntropy tends to the covariance analogously to the conventional one.

3. Maximum Complex Correntropy Criterion

A typical system identification task is represented in Fig. 1. Since correntropy has been previously defined in the complex domain, it is necessary to establish the MCCC. Let the new cost function J_{MCCC} be the maximum complex correntropy between two random complex variables, where D is the desired signal and Y is the filter output, while D, Y are complex-valued random variables.

$$J_{MCCC} = \max_w V(D, Y) \quad (16)$$

The filter output consists in the combination of the system input x and weights w , where $x, w \in \mathbb{C}$. Then, it gives:

$$\begin{aligned}
Y &= wX \\
w &= w^{re} + jw^{im}
\end{aligned} \quad (17)$$

and

$$\begin{aligned}
D &= D^{re} + jD^{im} \\
X &= X^{re} + jX^{im}
\end{aligned} \quad (18)$$

In order to use correntropy in this case, let us assume $C_1 = D$ and $C_2 = Y$ in equation (13). Then x_n becomes

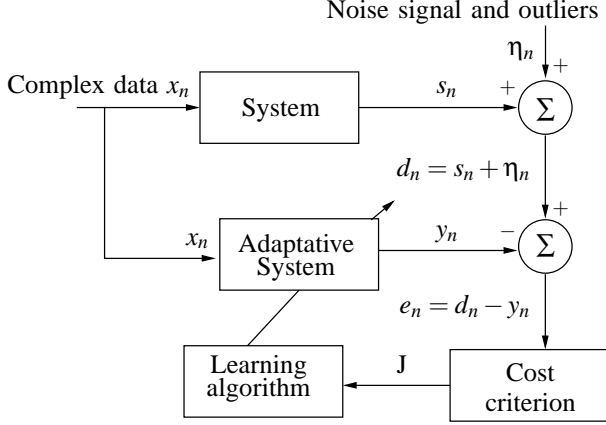


Figure 1: Typical system identification

d_n^{re} , y_n corresponds to $w_n^{re} x_n^{re} - w_n^{im} x_n^{im}$, and so on. The correntropy-based cost function can be expressed as:

$$J_{MCCC} = \frac{1}{N} \sum_{n=1}^N G_{\sigma\sqrt{2}}(d_n^{re} - (w_n^{re} x_n^{re} - w_n^{im} x_n^{im})) G_{\sigma\sqrt{2}}(d_n^{im} - (w_n^{re} x_n^{im} + w_n^{im} x_n^{re})) \quad (19)$$

The fixed-point solution for the optimal weights can be obtained by setting the cost function derivative to zero in equation (19):

$$\frac{\partial J_{MCCC}}{\partial w^{re}} = 0 \quad \text{and} \quad \frac{\partial J_{MCCC}}{\partial w^{im}} = 0 \quad (20)$$

isolating w^{re} and w^{im} gives:

$$w^{re} = \frac{\sum_{n=1}^N g_1(d, w, x) g_2(d, w, x) (d_n^{re} x_n^{re} + d_n^{im} x_n^{im})}{\sum_{n=1}^N g_1(d, w, x) g_2(d, w, x) ((x_n^{re})^2 + (x_n^{im})^2)} \quad (21)$$

$$w^{im} = \frac{\sum_{n=1}^N g_1(d, w, x) g_2(d, w, x) (d_n^{im} x_n^{re} - d_n^{re} x_n^{im})}{\sum_{n=1}^N g_1(d, w, x) g_2(d, w, x) ((x_n^{re})^2 + (x_n^{im})^2)} \quad (22)$$

where

$$g_1(d, w, x) = G_{\sigma\sqrt{2}}(d_n^{re} - (w^{re} x_n^{re} - w^{im} x_n^{im})) \quad (23)$$

$$g_2(d, w, x) = G_{\sigma\sqrt{2}}(d_n^{im} - (w^{re} x_n^{im} + w^{im} x_n^{re})) \quad (24)$$

The result provided by equations (23) and (24) represents the iterative solution for w^{re} and w^{im} . Even though convergence is achieved after a few iterations,

each one of them requires the computation of the whole sum, which is inadequate to real-time learning. A fixed-point stochastic recursive solution can then be derived as inspired by [20] and based on equations (23) and (24). Firstly, let us define the iterations for (23) and (24) as:

$$w^{re}[n] = \frac{P[n]}{R[n]} \quad (25)$$

$$w^{im}[n] = \frac{Q[n]}{R[n]} \quad (26)$$

Applying a stochastic approach to the sum gives:

$$P[n] = P[n-1] + g_1(d, w, x) g_2(d, w, x) (d_{n-1}^{re} x_{n-1}^{re} + d_{n-1}^{im} x_{n-1}^{im}) \quad (27)$$

$$Q[n] = Q[n-1] + g_1(d, w, x) g_2(d, w, x) (d_{n-1}^{im} x_{n-1}^{re} - d_{n-1}^{re} x_{n-1}^{im}) \quad (28)$$

$$R[n] = R[n-1] + g_1(d, w, x) g_2(d, w, x) ((x_{n-1}^{re})^2 + (x_{n-1}^{im})^2) \quad (29)$$

In order to implement the recursive expressions represented by (27), (28), and (29), it is necessary to compute the initial parameters w^{re} and w^{im} , as well as compute the initial values for $P[0]$, $Q[0]$, $R[0]$.

4. Simulation and Results

In order to evaluate the MCCC performance, the complex RLS algorithm presented in [15] has been adopted for comparison purposes. Besides, the weight signal-to-noise ratio (WSRN) is also considered in the analysis of results as in [20], since it quantifies convergence and misadjustment rates properly in decibels as:

$$WSNR_{db} = 10 \log_{10} \left(\frac{\bar{w} \bar{w}^*}{(\bar{w} - w[n])(\bar{w} - w[n])^*} \right) \quad (30)$$

Where $\bar{w} = \bar{w}^{re} + j \bar{w}^{im}$ is the proper weight chosen for the simulation tests and $w[n] = w^{re}[n] + j w^{im}[n]$ is the weight computed by the aforementioned methods in the n -th iteration.

The desired signal is contaminated with nonGaussian noise whose PDF (probability density function) is $0.95\mathcal{N}(0.0, 0.05) + 0.05\mathcal{N}(0.0, 5.0)$, where $\mathcal{N}(\mu, \sigma)$ is a normal Gaussian distribution with mean μ and variance σ^2 . The authors in [20] also employ the aforementioned PDF to represent the noise and evaluate robustness of the fixed-point MCC algorithm compared

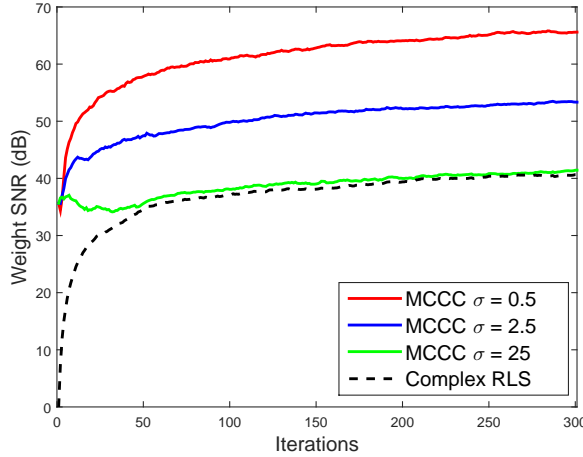


Figure 2: Weight SNR plots for MCC fixed point

with its RLS counterpart, although data only comprises the real domain. The noise signal η_n is then generated in this work, where $\eta \in \mathbb{C}$ and $\eta_n = \eta_n^{re} + j\eta_n^{im}$, while η_n^{re} and $\eta_n^{im} \in \mathbb{R}$ and follow the described PDF.

After 300 iterations, the results shown in Fig. 2 could be obtained. The curves represent the average when using 50 Monte Carlo trials, as the weights always start from random values. Fast convergence is achieved by both methods as in [20], but improved performance is achieved when the kernel size is $\sigma=0.5$. It can be stated that the proposed approach is able to ignore outliers. The kernel size in equation (14) behaves as a parameter that weights both second-order ($m=1$) and higher-order moments. As σ becomes higher than unity, the high-order moments decrease faster as the achieved results are closer to the ones provided by the conventional complex RLS solution.

5. Conclusions

This paper has presented the extension of the correntropy concept to complex-valued data in an approach defined as complex correntropy. A significant contribution of this work lies in obtaining the expression for the complex correntropy from its respective probabilistic interpretation. Besides, a recursive algorithm based on fixed-point solution has been introduced, which can be used to derive the MCCC. Simulation tests have also demonstrated that the proposed method presents high convergence rates, but with higher efficiency when dealing with outlier environments if compared to the complex RLS approach. It is then reasonable to state that correntropy can now be applied to the solution of distinct problems involving complex data in a more

straightforward way.

References

- [1] Z. Xu, Y. Ji, and D. Zhou, "Real-time reliability prediction for a dynamic system based on the hidden degradation process identification," *Reliability, IEEE Transactions on*, vol. 57, no. 2, pp. 230–242, June 2008.
- [2] T. Liu, K. Yao, and F. Gao, "Identification and autotuning of temperature-control system with application to injection molding," *Control Systems Technology, IEEE Transactions on*, vol. 17, no. 6, pp. 1282–1294, Nov 2009.
- [3] N. Karamolegkos, F. Vicario, and N. Chbat, "Cardio-vascular system identification: Simulation study using arterial and central venous pressures," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, Aug 2015, pp. 997–1000.
- [4] A. I. R. Fontes, P. T. V. Souza, A. D. D. Neto, A. de M. Martins, and L. F. Q. Silveira, "Classification system of pathological voices using correntropy," *Mathematical Problems in Engineering*, vol. 2014, no. 7, Aug 2014.
- [5] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, Nov 2007.
- [6] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *2009 International Joint Conference on Neural Networks*, June 2009, pp. 2950–2955.
- [7] A. I. Fontes, A. de M. Martins, L. F. Silveira, and J. Principe, "Performance evaluation of the correntropy coefficient in automatic modulation classification," *Expert Systems with Applications*, vol. 42, no. 1, pp. 1–8, 2015.
- [8] I. Santamaria, P. Pokharel, and J. Principe, "Generalized correlation function: definition, properties, and application to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.
- [9] J.-W. Xu and J. C. Principe, "A pitch detector based on a generalized correlation function," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 1420–1432, 2008.
- [10] L. L. Linhares, A. I. Fontes, A. M. Martins, F. M. Araújo, and L. F. Silveira, "Fuzzy wavelet neural network using a correntropy criterion for nonlinear system identification," *Mathematical Problems in Engineering*, vol. 2015, 2015.
- [11] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-gaussian signal processing."

- [12] Y. Liu and J. Chen, "Correntropy kernel learning for nonlinear system identification with outliers," *Industrial & Engineering Chemistry Research*, vol. 53, no. 13, pp. 5248–5260, 2013.
- [13] J. B. Rego, A. I. Fontes, A. a. D. Neto, L. F. Silveira, and A. M. Martins, "Linear regression based on correntropy for sensor calibration," *Simpósio Brasileiro de Automação Inteligente (SBAI)*, vol. 2, p. 2, 2015.
- [14] R. He, W. S. Zheng, and B. G. Hu, "Maximum correntropy criterion for robust face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1561–1576, 2011.
- [15] P. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, ser. Kluwer international series in engineering and computer science. Springer, 2008. [Online]. Available: <https://books.google.com.br/books?id=Bec7aJWljFIC>
- [16] N.-Y. Kim, "Complex-channel blind equalization using cross-correntropy," *Journal of Internet Computing and Services*, vol. 11, no. 5, pp. 19–26, 2010.
- [17] B. D. Azam Khalili, Amir Rastegarnia, "A robust adaptive carrier frequency offset estimation algorithm for ofdm."
- [18] J. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*, ser. Information Science and Statistics. Springer New York, 2010. [Online]. Available: <https://books.google.com.br/books?id=oJSkBXWctsgC>
- [19] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: A localized similarity measure," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 4919–4924.
- [20] A. Singh and J. C. Principe, "A closed form recursive solution for maximum correntropy training," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2070–2073.

This figure "Arquitetura.jpg" is available in "jpg" format from:

<http://arxiv.org/ps/1606.04761v1>